Full wwPDB X-ray Structure Validation Report (i)

Nov 3, 2016 – 02:15 PM EDT

PDB ID	:	5KHX
Title	:	Crystal structure of JAK1 in complex with PF-4950736
Authors	:	Han, S.; Caspers, N.L.
Deposited on	:	2016-06-16
Resolution	:	2.40 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

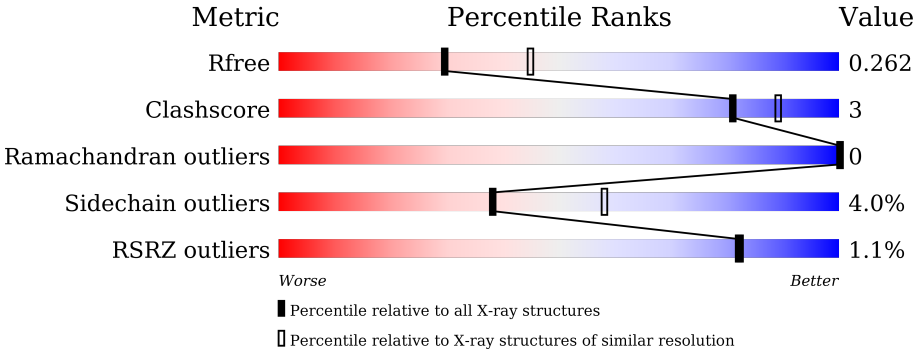
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028320
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028320

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

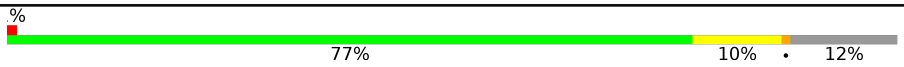
The reported resolution of this entry is 2.40 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	2919 (2.40-2.40)
Clashscore	102246	3407 (2.40-2.40)
Ramachandran outliers	100387	3351 (2.40-2.40)
Sidechain outliers	100360	3352 (2.40-2.40)
RSRZ outliers	91569	2928 (2.40-2.40)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	316	

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2387 atoms, of which 17 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

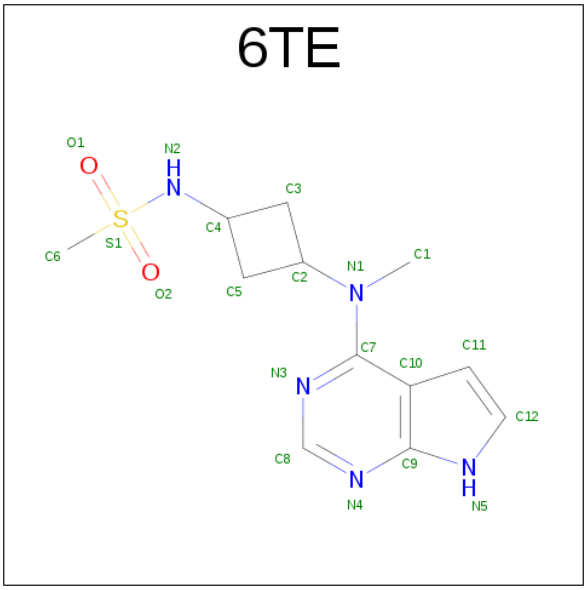
- Molecule 1 is a protein called Tyrosine-protein kinase JAK1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	P	S			
1	A	278	2271	1444	388	422	2	15	0	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	839	GLY	-	expression tag	UNP P23458
A	840	SER	-	expression tag	UNP P23458

- Molecule 2 is {N}-[3-[methyl(7 {H}-pyrrolo[2,3-d]pyrimidin-4-yl)amino]cyclobutyl]methane sulfonamide (three-letter code: 6TE) (formula: $C_{12}H_{17}N_5O_2S$).



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	H	N	O	S		
2	A	1	37	12	17	5	2	1	17	0

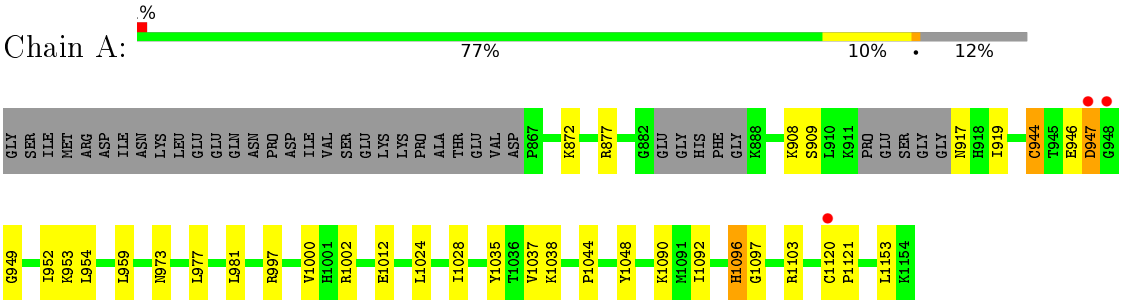
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	79	79	79	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: Tyrosine-protein kinase JAK1



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	46.25Å 88.30Å 146.84Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	73.42 – 2.40 73.42 – 2.40	Depositor EDS
% Data completeness (in resolution range)	96.8 (73.42-2.40) 97.0 (73.42-2.40)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.45 (at 2.40Å)	Xtriage
Refinement program	BUSTER 2.11.6	Depositor
R, R_{free}	0.170 , 0.242 0.175 , 0.262	Depositor DCC
R_{free} test set	561 reflections (5.00%)	DCC
Wilson B-factor (Å^2)	39.1	Xtriage
Anisotropy	0.308	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 63.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	2387	wwPDB-VP
Average B, all atoms (Å^2)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.94% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: 6TE, PTR

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| >5$	RMSZ	$\#\|Z\| >5$
1	A	0.49	0/2286	0.70	3/3073 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	2

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	1097	GLY	N-CA-C	-7.47	94.43	113.10
1	A	1096	HIS	N-CA-C	-5.85	95.21	111.00
1	A	1096	HIS	C-N-CA	5.24	133.30	122.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1096	HIS	Mainchain,Peptide

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within

the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2271	0	2274	15	0
2	A	20	17	0	0	0
3	A	79	0	0	0	0
All	All	2370	17	2274	15	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 3.

All (15) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:947:ASP:HA	1:A:949:GLY:H	1.30	0.97
1:A:947:ASP:HA	1:A:949:GLY:N	2.09	0.66
1:A:944:CYS:SG	1:A:953:LYS:HD3	2.38	0.63
1:A:872:LYS:HD2	1:A:946:GLU:HG2	1.88	0.56
1:A:1000:VAL:HG13	1:A:1028:ILE:HD11	1.90	0.54
1:A:981:LEU:HB2	1:A:1153:LEU:HD21	1.90	0.53
1:A:1002:ARG:CZ	1:A:1037:VAL:HG21	2.41	0.51
1:A:919:ILE:HD11	1:A:952:ILE:H	1.75	0.50
1:A:1035:PTR:HE1	1:A:1037:VAL:HG22	1.95	0.48
1:A:1044:PRO:HD2	1:A:1048:TYR:CZ	2.52	0.44
1:A:959:LEU:HD13	1:A:1012:GLU:HA	1.98	0.44
1:A:908:LYS:HB3	1:A:954:LEU:HB3	1.99	0.43
1:A:977:LEU:HD21	1:A:1121:PRO:HG3	2.01	0.42
1:A:1092:ILE:HG23	1:A:1103:ARG:HB3	2.03	0.41
1:A:997:ARG:HA	1:A:997:ARG:HD3	1.89	0.41

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	271/316 (86%)	263 (97%)	8 (3%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	250/281 (89%)	240 (96%)	10 (4%)	38	58

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	877	ARG
1	A	909	SER
1	A	917	ASN
1	A	944	CYS
1	A	947	ASP
1	A	973	ASN
1	A	1024	LEU
1	A	1038	LYS
1	A	1090	LYS
1	A	1120	CYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	936	ASN
1	A	971	ASN
1	A	1072	HIS

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

2 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
1	PTR	A	1034	1	13,16,17	1.55	3 (23%)	19,22,24	2.06	8 (42%)
1	PTR	A	1035	1	13,16,17	1.35	1 (7%)	19,22,24	1.66	6 (31%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
1	PTR	A	1034	1	-	0/9/11/13	0/1/1/1
1	PTR	A	1035	1	-	0/9/11/13	0/1/1/1

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	1034	PTR	CE2-CZ	2.19	1.43	1.38
1	A	1034	PTR	CD2-CG	2.43	1.44	1.38
1	A	1034	PTR	CD1-CG	2.99	1.45	1.38
1	A	1035	PTR	CE2-CZ	3.32	1.45	1.38

All (14) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	1034	PTR	OH-CZ-CE2	-3.35	108.79	119.22
1	A	1035	PTR	CG-CB-CA	-2.65	108.02	114.12
1	A	1034	PTR	CE2-CD2-CG	-2.41	117.64	121.02
1	A	1035	PTR	CE2-CD2-CG	-2.32	117.76	121.02
1	A	1034	PTR	CB-CG-CD2	-2.30	116.26	120.91
1	A	1035	PTR	O-C-CA	-2.04	120.25	125.72
1	A	1034	PTR	CD2-CG-CD1	2.03	121.50	118.15

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	1035	PTR	O2P-P-OH	2.15	112.31	105.47
1	A	1035	PTR	OH-CZ-CE2	2.42	126.74	119.22
1	A	1034	PTR	O3P-P-O2P	2.79	117.67	107.44
1	A	1034	PTR	OH-CZ-CE1	2.93	128.33	119.22
1	A	1034	PTR	P-OH-CZ	3.55	133.93	123.85
1	A	1035	PTR	O3P-P-OH	3.63	117.02	105.47
1	A	1034	PTR	O3P-P-OH	3.64	117.06	105.47

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	1035	PTR	1	0

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	6TE	A	4000	-	19,22,22	1.51	3 (15%)	18,33,33	1.12	1 (5%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	6TE	A	4000	-	-	0/7/21/21	0/3/3/3

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	4000	6TE	C6-S1	-2.83	1.67	1.75
2	A	4000	6TE	C7-N3	2.85	1.38	1.34
2	A	4000	6TE	C9-N4	3.76	1.38	1.34

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	4000	6TE	C8-N3-C7	2.33	117.14	111.64

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	276/316 (87%)	-0.40	3 (1%) 82 82	24, 47, 81, 107	0

All (3) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	947	ASP	4.7
1	A	948	GLY	4.5
1	A	1120	CYS	2.1

6.2 Non-standard residues in protein, DNA, RNA chains (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
1	PTR	A	1035	16/17	0.93	0.14	-	32,53,73,75	0
1	PTR	A	1034	16/17	0.93	0.12	-	39,47,66,66	0

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron

density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
2	6TE	A	4000	20/20	0.98	0.13	-0.13	31,37,41,44	17

6.5 Other polymers (i)

There are no such residues in this entry.